FAN HAI-FU (1965). *Acta Phys. Sin.* **21**, 1105-1113.
GIACOVAZZO, C. (1983). *Acta Cryst.* A**39**, 685-692.
GOULD, R. O., VAN DEN HARK, TH. E. M. & BEURSKENS, P. T. (1975). *Acta Cryst.* A**31**, 813-817.
HEINERMAN, J. J. L. (1977). *Acta Cryst.* A**33**, 100-106.
KARLE, J. & KARLE, I. (1966). *Acta Cryst.* **21**, 849-859.
KELLER-SCHIERLEIN, W., MEYER, M., CELLAI, L., CERRINI, S., LAMBA, D., SEGRE, A., FEDELI, W. & BRUFANI, M. (1985). *J. Antibiot.* In the press.
KRABBENDAM, H. & KROON, J. (1971). *Acta Cryst.* A**27**, 48-53.
MAIN, P. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 97-105. Copenhagen: Munksgaard.
MAIN, P. (1978). *Acta Cryst.* A**34**, 31-38.
NUNZI, A., BURLA, M. C., POLIDORI, G., GIACOVAZZO, C., CASCARANO, G., VITERBO, D., CAMALLI, M. & SPAGNA, R. (1984). *Acta Cryst.* A**40**, C425.
PRICK, P. A. J., BEURSKENS, P. I. & GOULD, R. O. (1983). *Acta Cryst.* A**39**, 570-576.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60-65.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813-815.

Acta Cryst. (1985). A**41**, 613-617

New Methods for Deriving Joint Probability Distributions of Structure Factors. I

BY J. BROSIUS*

Katholieke Universiteit Leuven, Departement Wiskunde, Belgium

(*Received* 23 *January* 1984; *accepted* 3 *June* 1985)

Abstract

With new probabilities, based on the Patterson function, for the 'atomic' random variables $\mathbf{x}_1, \ldots, \mathbf{x}_N$ in $P\bar{1}$, it is shown that an improved estimate can be obtained for the sign of the seminvariant $E_{2\mathbf{h}}$ in $P\bar{1}$. Two probability measures are considered. A method is also given for the case of a known Patterson vector of the form $2\mathbf{r}_1$, giving an estimate for the sign of any structure factor $E_\mathbf{h}$ by using its first neighborhood.

1. Introduction

For deriving joint probability distributions of structure factors one has used up to now two conceptually different approaches. One is to consider the structure factor

$$E_\mathbf{h} = \left(\sum_{i=1}^{N} f_i^2\right)^{-1/2} \sum_{i=1}^{N} f_i \exp(2\pi i \mathbf{h}\,.\,\mathbf{x}_i) \qquad (1)$$

as a function of the random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$; the other consists in regarding $E_\mathbf{h}$ as a function of the random variable $\mathbf{h}$. The first method consists in letting the random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ range uniformly and independently over the unit cell, which may be represented mathematically by $[0, 1[^3$ [the set of all triples (u, v, w) where $0 \le u, v, w < 1$]. In this paper other probability measures are considered for the random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ based on the Patterson function. In particular, we study the seminvariant $E_{2\mathbf{h}}$ in $P\bar{1}$.

* Present address: Université du Burundi, Département de Máthematiques, BP 2700 Bujumbura, Burundi.

2. The probability distribution of $E_{2\mathbf{h}}$ in $P\bar{1}$ for different probabilities for $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$

Several probability measures for $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ will be considered and used to determine the sign of $E_{2\mathbf{h}}$ for its first neighborhood. In order to simplify calculations we shall treat the case of N equal atoms for which the structure factor $E_\mathbf{h}$ is given by

$$E_\mathbf{h} = 2N^{-1/2} \sum_{i=1}^{t} \cos(2\pi \mathbf{r}_i\,.\,\mathbf{h})$$

($\mathbf{r}_i \in [0, 1[^3$ and $t = N/2$). The function Q defined on $[0, 1[^3$ by

$$\mathbf{u} \in [0, 1[^3 \rightarrow Q(\mathbf{u}) = \langle (E_\mathbf{k}^2 - 1) \exp(-2\pi i \mathbf{k}\,.\,\mathbf{u}) \rangle_\mathbf{k} \qquad (2)$$

(where $\langle\,.\,\rangle_\mathbf{k}$ means the average over all reciprocal-lattice vectors) gives

$$Q(\mathbf{u}) = \begin{cases} N^{-1} & \text{if } \mathbf{u} = [2\mathbf{r}_i] \text{ or } \mathbf{u} = [-2\mathbf{r}_i] \quad (1 \le i \le t) \\ 2N^{-1} & \text{if } \mathbf{u} = [\mathbf{r}_i - \mathbf{r}_j] \text{ or } \mathbf{u} = [\mathbf{r}_i + \mathbf{r}_j] \text{ or} \\ & \quad \mathbf{u} = [-\mathbf{r}_i - \mathbf{r}_j] \quad (1 \le i, j \le t \text{ and } i \ne j) \\ 0 & \text{elsewhere,} \end{cases} \qquad (3)$$

where $[\mathbf{x}]$ for $\mathbf{x} \in \mathbb{R}^3$ denotes the unique vector in $[0, 1[^3$, which differs from $\mathbf{x}$ by some vector (p, q, r), where p, q and r are integer numbers.

This function Q will be used to construct several probability measures on the 'atomic' random variables $\mathbf{x}_i (1 \le i \le t)$. The simplest probability measure is obtained as follows. The random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ will be taken to be independent. They are defined on $[0, 1[^3$, equipped with its usual collection of Borel sets, by $\mathbf{u} \in [0, 1[^3 \rightarrow \mathbf{x}_i(\mathbf{u}) = \mathbf{u}$ $(1 \le i \le t)$.

0108-7673/85/060613-05$01.50

We shall now define a probability on the set $[0, 1[^3$. For any positive integer n let A_n denote the set of all reciprocal-lattice vectors $\mathbf{h}=(h_1, h_2, h_3)$ such that $-n\le h_i\le n$ $(1\le i\le 3)$. Define the probability P on $[0, 1[^3$ by

$$P(B)=\lim_{n\to+\infty}\int_B\left[\sum_{\mathbf{k}\in A_n}\frac{E_{\mathbf{k}}^2-1}{N-1}\exp(-2\pi i(2\mathbf{u})\,.\,\mathbf{k})\right]\mathrm{d}\mathbf{u}, \quad (4)$$

where B is any Borel set in $[0, 1[^3$ {in particular, where B is a set of the form $[a_1, b_1]\times[a_2, b_2]\times[a_3, b_3]$, which is the set of all triples (u_1, u_2, u_3) such that $a_i\le u_i\le b_i$ $(i=1,2,3)$ and with $0\le a_i\le b_i<1$ $(i=1,2,3)$}. It may be noted that P is a convex sum of point measures {that is P is of the form $P=\sum_{\mathbf{a}}\lambda_{\mathbf{a}}\delta_{\mathbf{a}}$, where the summation is over the finite set of Patterson vectors $\mathbf{a}$ in $[0, 1[^3$, $0<\lambda_{\mathbf{a}}<1$ and $\sum_{\mathbf{a}}\lambda_{\mathbf{a}}=1$, and where $\delta_{\mathbf{a}}$ denotes the point measure (or Dirac measure) in $\mathbf{a}$ with total mass equal to 1}. As usual we shall use the symbol $[\mathbf{x}_i\in B]$ to denote the event that $\mathbf{x}_i\in B$ (more precisely that $\mathbf{x}_i$ will take its values in B). We then get for the probability, $P([\mathbf{x}_i\in B])$ (where B is a Borel set in $[0, 1[^3$), that $\mathbf{x}_i\in B$: $P([\mathbf{x}_i\in B])=P(B)$ and for the mean, $\varepsilon[\cos(2\pi\mathbf{x}_i\,.\,\mathbf{h})]$, of the random variable $\cos(2\pi\mathbf{x}_i\,.\,\mathbf{h})$ for $1\le i\le t$

$$\begin{aligned}\varepsilon[\cos(2\pi\mathbf{x}_i\,.\,\mathbf{h})]&=\int\cos(2\pi\mathbf{u}\,.\,\mathbf{h})\,\mathrm{d}P(\mathbf{u})\\&=\lim_{n\to+\infty}\int\Big\{\sum_{\mathbf{k}\in A_n}[(E_{\mathbf{k}}^2-1)/(N-1)]\\&\quad\times\exp[-2\pi i(2\mathbf{u})\,.\,\mathbf{k}]\Big\}\cos(2\pi\mathbf{u}\,.\,\mathbf{h})\,\mathrm{d}\mathbf{u}.\end{aligned}\quad (5)$$

Hence

$$\varepsilon\{\cos[2\pi\mathbf{x}_i\,.\,(2\mathbf{h})]\}=(E_{\mathbf{h}}^2-1)/(N-1)\quad(1\le i\le t) \quad (6)$$

and $\varepsilon[\cos(2\pi\mathbf{x}_i\,.\,\mathbf{h})]=0$, if $\frac{1}{2}\mathbf{h}$ is not a reciprocal vector. Let us denote by $\hat{E}_{\mathbf{h}}$ the random variable

$$\hat{E}_{\mathbf{h}}=2N^{-1/2}\sum_{i=1}^{t}\cos(2\pi\mathbf{x}_i\,.\,\mathbf{h}) \quad (7)$$

and so we get

$$\varepsilon(\hat{E}_{2\mathbf{h}})=N^{1/2}(E_{\mathbf{h}}^2-1)/(N-1).$$

Let us denote by $P(E_{2\mathbf{h}})$ the density distribution of the random variable $\hat{E}_{2\mathbf{h}}$. We may develop $P(E_{2\mathbf{h}})$ into an asymptotic series according to powers of $N^{-1/2}$ (Bourbaki, 1961). We then get for the conditional probability, denoted as usual by $P_+(E_{2\mathbf{h}})$, that the sign of $\hat{E}_{2\mathbf{h}}$ is positive given $|\hat{E}_{2\mathbf{h}}|=|E_{2\mathbf{h}}|$ *up to the order* $N^{-1/2}$:

$$P_+(E_{2\mathbf{h}})=\tfrac{1}{2}+\tfrac{1}{2}\tanh[|E_{2\mathbf{h}}|(E_{\mathbf{h}}^2-1)N^{-1/2}] \quad (8)$$

(see Appendix).

It is interesting to note that

$$\sigma^2(\hat{E}_{2\mathbf{h}})=1+N^{-1}(E_{2\mathbf{h}}^2-1). \quad (9)$$

Hence, if $E_{2\mathbf{h}}^2$ is not too high the variance $\sigma^2(\hat{E}_{2\mathbf{h}})$ differs little from the variance $\sigma_u^2(\hat{E}_{2\mathbf{h}})$ of $\hat{E}_{2\mathbf{h}}$, which equals 1, where the index u in $\sigma_u^2(\hat{E}_{2\mathbf{h}})$ refers to the usual probability measure on $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t$ $(t=N/2)$.

Consider now the second neighborhood, $\{\hat{E}_{2\mathbf{h}}, \hat{E}_{\mathbf{h}}\}$, of $\hat{E}_{2\mathbf{h}}$ [for the notion of neighborhood see *e.g.* Hauptman (1976)]. Again let us denote by $P(E_{2\mathbf{h}}, E_{\mathbf{h}})$ the joint distribution of the random variables $\hat{E}_{2\mathbf{h}}$ and $\hat{E}_{\mathbf{h}}$. Developing $P(E_{2\mathbf{h}}, E_{\mathbf{h}})$ asymptotically according to powers of $N^{-1/2}$, we then get for the conditional probability, denoted by $P_+(E_{2\mathbf{h}}||E_{\mathbf{h}}|)$, that the sign of $\hat{E}_{2\mathbf{h}}$ is positive given $|\hat{E}_{2\mathbf{h}}|=|E_{2\mathbf{h}}|$ and $|\hat{E}_{\mathbf{h}}|=|E_{\mathbf{h}}|$ (assuming that $\frac{1}{2}\mathbf{h}$ is not a reciprocal vector) *up to the order* $N^{-1/2}$:

$$P_+(E_{2\mathbf{h}}||E_{\mathbf{h}}|)=\tfrac{1}{2}+\tfrac{1}{2}\tanh[|\tfrac{3}{2}|E_{2\mathbf{h}}|(E_{\mathbf{h}}^2-1)N^{-1/2}|]. \quad (10)$$

Again let us denote by $\sigma_u^2(\hat{E}_{2\mathbf{h}}||\hat{E}_{\mathbf{h}}|=|E_{\mathbf{h}}|)$ the conditional variance of $\hat{E}_{2\mathbf{h}}$ given $|\hat{E}_{\mathbf{h}}|=|E_{\mathbf{h}}|$ for the usual probability measure on $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t$. Then we find, up to the order N^{-1},

$$\begin{aligned}\sigma^2(\hat{E}_{2\mathbf{h}}||\hat{E}_{\mathbf{h}}|=|E_{\mathbf{h}}|)&=\sigma_u^2(\hat{E}_{2\mathbf{h}}||\hat{E}_{\mathbf{h}}|=|E_{\mathbf{h}}|)\\&\quad+N^{-1}[(E_{2\mathbf{h}}^2-1)-(E_{\mathbf{h}}^2-1)^2]\end{aligned} \quad (11)$$

(see Appendix). So, if $E_{2\mathbf{h}}^2-1\le(E_{\mathbf{h}}^2-1)^2$ we find that (up to order N^{-1})

$$\sigma^2(\hat{E}_{2\mathbf{h}}||\hat{E}_{\mathbf{h}}|=|E_{\mathbf{h}}|)\le\sigma_u^2(\hat{E}_{2\mathbf{h}}||\hat{E}_{\mathbf{h}}|=|E_{\mathbf{h}}|).$$

Also, note that it is always true that $\sigma^2(\hat{E}_{2\mathbf{h}}|\hat{E}_{\mathbf{h}})\le\sigma^2(\hat{E}_{2\mathbf{h}})$ (Barra, 1971).

These results differ significantly from the well known results (Klug, 1948) obtained by letting the $\mathbf{x}_i$ range uniformly over the whole unit cell. Consider now the function Q' on $[0, 1[^3$ defined by

$$\mathbf{u}\in[0, 1[^3\to Q'(\mathbf{u})=\sum_{\mathbf{k}}(E_{\mathbf{k}}^2-1)\exp(-2\pi i\mathbf{u}\,.\,\mathbf{k}), \quad (12)$$

where the summation is over a finite subset (the measured values) of the reciprocal lattice. If there are enough terms in (12) then Q' is positive almost everywhere. A density function of the $\mathbf{x}_i$ $(1\le i\le t)$ for which the $\mathbf{x}_i$ are no longer independent random variables is a density function proportional to

$$\left[\prod_{\substack{i=1\\i<j}}^{t}\prod_{j=1}^{t}Q'(\mathbf{x}_i-\mathbf{x}_j)Q'(\mathbf{x}_i+\mathbf{x}_j)\right]\prod_{i=1}^{t}Q'(2\mathbf{x}_i), \quad (13)$$

where in (13) we have used (by abuse of notation) the symbol $\mathbf{x}_i$ also for the argument in Q'. However, (13) does not allow an asymptotic development in powers of $N^{-1/2}$ to calculate the joint density distribution of a set of structure factors and we also get multiple averages over the given subset of the reciprocal lattice if one calculates the mean $\varepsilon[\cos 2\pi\mathbf{h}\,.\,(2\mathbf{x}_i)]$. So (13) is not suitable from a practical point of view. However, we may still use the asymptotic development to calculate the joint density distribution if we consider the following density function for the variables $\mathbf{x}_i$. For the sake of simplicity suppose that t ($t=N/2$) is a multiple of 2. We may then arrange the variables $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t\}$ in groups of two variables $\{(\mathbf{x}_1, \mathbf{x}_2), (\mathbf{x}_3, \mathbf{x}_4), \dots, (\mathbf{x}_{t-1}, \mathbf{x}_t)\}$ and use as density a function proportional to

$$Q'(2\mathbf{x}_1)Q'(2\mathbf{x}_2)Q'(\mathbf{x}_1-\mathbf{x}_2) \qquad (14)$$

for any such group [here the group $(\mathbf{x}_1, \mathbf{x}_2)$ of two variables]. The total density function is then proportional to

$$\prod_{i=0}^{t/2-1} [Q'(2\mathbf{x}_{2i+1})Q'(2\mathbf{x}_{2i+2})Q'(\mathbf{x}_{2i+1}-\mathbf{x}_{2i+2})]. \qquad (15)$$

Then we get for the mean $\varepsilon(\hat{E}_{2\mathbf{h}})$ of $\hat{E}_{2\mathbf{h}}$

$$\varepsilon(\hat{E}_{2\mathbf{h}}) = 2N^{-1/2}\sum_{i=1}^{t} \varepsilon[\cos 2\pi\mathbf{x}_i\,.\,(2\mathbf{h})]$$
$$= N^{1/2}\frac{\langle(E_{\mathbf{k}}^2-1)(E_{2\mathbf{k}}^2-1)(E_{\mathbf{h}+\mathbf{k}}^2-1)\rangle_{\mathbf{k}}}{\langle(E_{\mathbf{k}}^2-1)^2(E_{2\mathbf{k}}^2-1)\rangle_{\mathbf{k}}}, \qquad (16)$$

where $\langle\,.\,\rangle_{\mathbf{k}}$ means the average over the given subset of the reciprocal lattice. Again we may calculate the density distribution $P(E_{2\mathbf{h}})$ for this new probability with a density given by (15) by using an asymptotic development in powers of $N^{-1/2}$. The main term of $P(E_{2\mathbf{h}})$ is also calculated by observing that $\hat{E}_{2\mathbf{h}}$ is normally distributed if t is high enough. We then get for the probability, denoted by $P_+(E_{2\mathbf{h}})$, that the sign of $\hat{E}_{2\mathbf{h}}$ is positive given $|\hat{E}_{2\mathbf{h}}|=|E_{2\mathbf{h}}|$:

$$P_+(E_{2\mathbf{h}}) = \frac{1}{2}+\frac{1}{2}\tanh\left[\frac{|E_{2\mathbf{h}}|N^{1/2}}{\sigma^2}\times\frac{\langle(E_{\mathbf{k}}^2-1)(E_{2\mathbf{k}}^2-1)(E_{h+k}^2-1)\rangle_k}{\langle(E_{\mathbf{k}}^2-1)^2(E_{2\mathbf{k}}^2-1)\rangle_{\mathbf{k}}}\right], \qquad (17)$$

where $\sigma^2=\sigma^2(\hat{E}_{2\mathbf{h}})=\varepsilon(\hat{E}_{2\mathbf{h}}^2)-\varepsilon(\hat{E}_{2\mathbf{h}})^2$ is given by

$$\sigma^2 \simeq 1+N^{-1/2}\varepsilon(\hat{E}_{4\mathbf{h}})-(2/N)\varepsilon(\hat{E}_{2\mathbf{h}})^2$$
$$\simeq 1+\frac{\langle(E_{\mathbf{k}}^2-1)(E_{2\mathbf{k}}^2-1)(E_{2\mathbf{h}+\mathbf{k}}^2-1)\rangle_{\mathbf{k}}}{\langle(E_{\mathbf{k}}^2-1)^2(E_{2\mathbf{k}}^2-1)\rangle_{\mathbf{k}}}$$
$$-2\left[\frac{\langle(E_{\mathbf{k}}^2-1)(E_{2\mathbf{k}}^2-1)(E_{\mathbf{h}+\mathbf{k}}^2-1)\rangle_{\mathbf{k}}}{\langle(E_{\mathbf{k}}^2-1)^2(E_{2\mathbf{k}}^2-1)\rangle_{\mathbf{k}}}\right]^2. \qquad (18)$$

The averages occurring in (17) and (18) can be calculated *in the case of no Patterson overlap* either directly or by considering the random variables $E_{\mathbf{k}}$, $E_{2\mathbf{k}}$, $E_{\mathbf{h}+\mathbf{k}}$ as functions of the random variable $\mathbf{k}$ [the approach of Hauptman & Karle (1958)] and by calculating their joint density distribution $P(E_{\mathbf{k}}, E_{2\mathbf{k}}, E_{\mathbf{h}+\mathbf{k}})$. We then obtain

$$N^{1/2}\frac{\langle(E_{\mathbf{k}}^2-1)(E_{2\mathbf{k}}^2-1)(E_{\mathbf{h}+\mathbf{k}}^2-1)\rangle_{\mathbf{k}}}{\langle(E_{\mathbf{k}}^2-1)^2(E_{2\mathbf{k}}^2-1)\rangle_{\mathbf{k}}}$$
$$\simeq \tfrac{1}{3}E_{2\mathbf{h}}+(\tfrac{2}{3}N^{-1/2})(E_{\mathbf{h}}^2-1)+O(N^{-3/2}). \qquad (19)$$

Reconsider now σ^2 in (18). We see that if $\varepsilon(\hat{E}_{4\mathbf{h}})$ is negative and large in absolute value, $\sigma^2(\hat{E}_{2\mathbf{h}})$ $(=\sigma^2)$ becomes very small. Note that in (17) a so-called renormalization term [the term $\langle(E_{\mathbf{k}}^2-1)^2(E_{2\mathbf{k}}^2-1)\rangle_{\mathbf{k}}$] appears in a rigorous way. This term is positive and equals in the *case of no Patterson overlap*

$$\langle(E_{\mathbf{k}}^2-1)^2(E_{2\mathbf{k}}^2-1)\rangle_{\mathbf{k}} \simeq N^{-1}(6-5N^{-1}) \qquad (20)$$

(see Appendix).

We may obtain better estimates of the sign of $\hat{E}_{2\mathbf{h}}$ if we consider higher neighborhoods, *e.g.* the neighborhood $\{\hat{E}_{2\mathbf{h}}, \hat{E}_{\mathbf{h}}\}$. Indeed, (17) has been calculated by considering only the first neighborhood $\{\hat{E}_{2\mathbf{h}}\}$ of $\hat{E}_{2\mathbf{h}}$. Also, $P_+(E_{2\mathbf{h}})$ in (17) should not be confused with the probability that the sign of $\hat{E}_{2\mathbf{h}}$ is positive given $|\hat{E}_{2\mathbf{h}}|=|E_{2\mathbf{h}}|$ and given $|\hat{E}_{\mathbf{k}}|=|E_k|$ for all $\mathbf{k}$. This result might be compared with that in the paper of Giacovazzo (1976) where a probabilistic treatment seemed to be given of the $B_{3,0}$ formula. (Hauptman & Karle, 1958) and the $\sum_1$ formula. However, in Giacovazzo's derivation no valid argument is given for the neglect of all multiple averages over the reciprocal lattice in the joint distribution of structure factors. In contrast, $\varepsilon(\hat{E}_{2\mathbf{h}})$ in (16) and $\sigma^2(\hat{E}_{2\mathbf{h}})$ in (18) in the present work have been derived rigorously. By abuse of notation we have also used the same symbol $P_+(E_{2\mathbf{h}})$ in (17) and (8), although they are derived from different probability laws for the variables $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t$. Up to now we have considered probability laws for the $\mathbf{x}_i$ in which $2\mathbf{x}_i$ (and $\mathbf{x}_i-\mathbf{x}_j$ and $\mathbf{x}_i+\mathbf{x}_j$) do not range uniformly over the set of Patterson vectors. A probability law on the $\mathbf{x}_i$ in which $2\mathbf{x}_i$ (and $\mathbf{x}_i-\mathbf{x}_j$ and $\mathbf{x}_i+\mathbf{x}_j$) range uniformly over the set of Patterson vectors can be constructed as follows. Let M be the lowest peak in Q' that may be associated with a Patterson vector. Then we might use for each $\mathbf{x}_i$ $(1\le i\le t)$ a density function proportional to

$$\min[Q'(2\mathbf{x}_i), M] \qquad (21)$$

or more complicated functions. But now all means of the form $\varepsilon[\cos(2\pi\mathbf{x}_i\,.\,2\mathbf{h})]$ have to be calculated numerically. For the density function given by (21) we then get for the probability that the sign of $\hat{E}_{2\mathbf{h}}$ is positive given $|\hat{E}_{2\mathbf{h}}|=|E_{2\mathbf{h}}|$:

$$P_+(E_{2\mathbf{h}}) = \tfrac{1}{2}+\tfrac{1}{2}\tanh[|E_{2\mathbf{h}}|\varepsilon(\hat{E}_{2\mathbf{h}})/\sigma^2(\hat{E}_{2\mathbf{h}})], \qquad (22)$$

where $\varepsilon(\hat{E}_{2\mathbf{h}})$ and $\sigma^2(\hat{E}_{2\mathbf{h}})$ are calculated numerically from the density function on $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t$ obtained from (21). It is interesting to note that by observing the peak height in Q [(2) and (3)] we can filter out the Patterson vectors of the form $2\mathbf{r}_i$ from the others

($\mathbf{r}_i+\mathbf{r}_j$ and $\mathbf{r}_i-\mathbf{r}_j$). But this can only be done with no Patterson overlap. Indeed, consider the $B_{2,0}$ formula (Cochran, 1954; Hauptman & Karle, 1958).

$$E_{2\mathbf{h}}=N^{1/2}[2(E_{\mathbf{h}}^2-1)-N\langle(E_{\mathbf{k}}^2-1)(E_{\mathbf{h}+\mathbf{k}}^2-1)\rangle_{\mathbf{k}}]. \tag{23}$$

One can easily construct a probability law for $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t$ such that $\varepsilon(\hat{E}_{2\mathbf{h}})$ gives the right side of (23) (Brosius, 1979). But this probability law no longer remains positive (and thus a probability law), even with mild Patterson overlap. Indeed let P_1 be the probability law of $\mathbf{x}_1$ (as defined by Brosius, 1979) and consider the random variables $\mathbf{x}_1, \ldots, \mathbf{x}_t$ to be independent and all having the same probability law as $\mathbf{x}_1$. But then one has clearly

$$P\{(\mathbf{x}_1\in[0,1[^3)\}=\int \mathrm{d}P_1=2(N-1)-N\langle(E_{\mathbf{k}}^2-1)^2\rangle_{\mathbf{k}}. \tag{24}$$

So that, since $\int \mathrm{d}P_1=1$ (the event $[\mathbf{x}_1\in[0,1[^3]$ is a true event), one has

$$2(N-1)-N\langle(E_{\mathbf{k}}^2-1)^2\rangle_{\mathbf{k}}=1. \tag{25}$$

In the case of *no Patterson overlap*

$$\langle(E_{\mathbf{k}}^2-1)^2\rangle_{\mathbf{k}}=2-3N^{-1}. \tag{26}$$

Substitution of (26) in (25) means that the left-hand side of (25) then indeed gives (1). But even with mild Patterson overlap $\langle(E_{\mathbf{k}}^2-1)^2\rangle_{\mathbf{k}}$ increases (Hauptman, 1964) and becomes rapidly greater than $2-2N^{-1}$ so that $\int \mathrm{d}P_1$ even becomes <0, which is absurd. But note that the probability law derived from (15) remains a probability law even with Patterson overlap. If a Patterson vector of the form $2\mathbf{r}_i$, say $2\mathbf{r}_1$, is known a probability law can be constructed for which $\varepsilon(\hat{E}_{\mathbf{h}})=E_{\mathbf{h}}$ (for every $\mathbf{h}$) in the *absence* of Patterson overlap.

3. The case of known Patterson vectors for $P\bar{1}$

Suppose one knows a Patterson vector of the form $2\mathbf{r}_1$. Then we may apply the Patterson superposition technique (Buerger, 1951). Indeed, theoretically (that is without Patterson overlap) one expects that the function $\mathbf{u}\to p(\mathbf{u})p(\mathbf{u}+2\mathbf{r}_1)$ or $\min\{p(\mathbf{u}), p(\mathbf{u}+2\mathbf{r}_i)\}$ with

$$p(\mathbf{u})=\langle(E_{\mathbf{k}}^2-1)\exp(-2\pi i\mathbf{u}\,.\,\mathbf{k})\rangle_{\mathbf{k}} \tag{27}$$

will give an image of the real structure. So let us use as density function of $\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_t$ the function

$$\mu(\mathbf{x}_2,\ldots,\mathbf{x}_t)=\prod_{i=2}^{t}\mu_i(\mathbf{x}_i), \tag{28}$$

where $\mu_i(\mathbf{x}_i)$ is proportional to

$$\left\{\sum_{\mathbf{k}}(E_{\mathbf{k}}^2-1)\exp[-2\pi i\mathbf{k}\,.\,(\mathbf{x}_i+\mathbf{r}_1)]\right\}\times\left\{\sum_{\mathbf{k}}(E_{\mathbf{k}}^2-1)\exp[-2\pi i\mathbf{k}\,.\,(\mathbf{x}_i-\mathbf{r}_1)]\right\}, \tag{29}$$

where the summation is over the finite set of measured $|E_{\mathbf{k}}|$ values and where by abuse of notation the same symbol $\mathbf{x}_i$ is used to represent the argument in (29). Then one can verify that for the mean $\varepsilon(\hat{E}_{\mathbf{h}})$ (for any $\mathbf{h}$) one gets

$$\varepsilon(\hat{E}_{\mathbf{h}})=2N^{-1/2}\cos 2\pi\mathbf{r}_1\,.\,\mathbf{h}+(N-2)N^{-1/2}\times\frac{\langle(E_{\mathbf{k}}^2-1)(E_{\mathbf{h}+\mathbf{k}}^2-1)\cos[2\pi\mathbf{r}_1\,.\,(\mathbf{h}+2\mathbf{k})]\rangle_{\mathbf{k}}}{\langle(E_{\mathbf{k}}^2-1)^2\cos 2\pi\mathbf{r}_1\,.\,(2\mathbf{k})\rangle_{\mathbf{k}}}. \tag{30}$$

In the case of no Patterson overlap the right-hand side of (30) is $E_{\mathbf{h}}$. In this way we can get the density distribution of $\hat{E}_{\mathbf{h}}$ (the first neighborhood of $\hat{E}_{\mathbf{h}}$) and one can calculate, using the asymptotic development, the probability that the sign of $\hat{E}_{\mathbf{h}}$ is positive given $|\hat{E}_{\mathbf{h}}|=|E_{\mathbf{h}}|$. A similar expression to that in (30) can be found in Heinerman, Krabbendam & Kroon (1975).

In a future publication the space group $P1$ will be dealt with.

I am very grateful to Dr H. Hauptman for his many valuable comments.

APPENDIX

1. Derivation of (8) and (9)

The joint distribution $P(E)$ of $\hat{E}_{2\mathbf{h}}$ is normally distributed for N high enough. So

$$P(E)\propto\exp\{-[E-\varepsilon(\hat{E}_{2\mathbf{h}})]^2/2\sigma^2(\hat{E}_{2\mathbf{h}})\}. \tag{A.1}$$

Also

$$\begin{aligned}\varepsilon(\hat{E}_{2\mathbf{h}})&=2N^{-1/2}\sum_{i=1}^{t}\varepsilon\{\cos[2\pi\mathbf{x}_i\,.\,(2\mathbf{h})]\}\\&=N^{1/2}(E_{\mathbf{h}}^2-1)/(N-1)\simeq(E_{\mathbf{h}}^2-1)N^{-1/2}\end{aligned} \tag{A.2}$$

and

$$\begin{aligned}\sigma^2(\hat{E}_{2\mathbf{h}})&=\varepsilon(\hat{E}_{2\mathbf{h}}^2)-\varepsilon(\hat{E}_{2\mathbf{h}})^2\\&=\varepsilon[1+N^{-1/2}\hat{E}_{4\mathbf{h}}+4N^{-1}\sum_{i}\sum_{\substack{j\\i\neq j}}\cos(2\pi\mathbf{x}_i\,.\,2\mathbf{h})\\&\quad\times\cos(2\pi\mathbf{x}_j\,.\,2\mathbf{h})]-[\varepsilon(\hat{E}_{2\mathbf{h}})]^2\\&=1+N^{-1/2}\varepsilon(\hat{E}_{4\mathbf{h}})\\&\quad+4N^{-1}\sum_{i}\sum_{\substack{j\\i\neq j}}\varepsilon(\cos(2\pi\mathbf{x}_i\,.\,2\mathbf{h}))\\&\quad\times\varepsilon(\cos 2\pi(\mathbf{x}_j\,.\,2\mathbf{h}))-[\varepsilon(\hat{E}_{2\mathbf{h}})]^2\\&=1+(E_{2\mathbf{h}}^2-1)/(N-1)\\&\quad+4N^{-1}(N/2)(N/2-1)\\&\quad\times(E_{\mathbf{h}}^2-1)^2/(N-1)^2\\&\quad-N(E_{\mathbf{h}}^2-1)^2/(N-1)^2\\&\simeq1+(E_{2\mathbf{h}}^2-1)/(N-1).\end{aligned} \tag{A.3}$$

So

$$P(E) \propto \exp\{[E - N^{-1/2}(E_{\mathbf{h}}^2-1)]^2 \times \tfrac{1}{2}[1+(E_{2\mathbf{h}}^2-1)/(N-1)]^{-1}\}. \quad (A.4)$$

Hence it follows that

$$P_+(E_{2\mathbf{h}}) = \tfrac{1}{2}+\tfrac{1}{2}\tanh\{|E_{2\mathbf{h}}|(E_{\mathbf{h}}^2-1)N^{-1/2} \times [1+(E_{2\mathbf{h}}^2-1)/(N-1)]^{-1}\} \simeq \tfrac{1}{2}+\tfrac{1}{2}\tanh[|E_{2\mathbf{h}}|(E_{\mathbf{h}}^2-1)N^{-1/2}]. \quad (A.5)$$

2. Derivation of (10) and (11)

Calculate the joint density $P(E_1, E_2)$, where $E_1 = E_{2\mathbf{h}}$ and $E_2 = E_{\mathbf{h}}$.

$$P(E_1, E_2) \propto \int \exp(-iuE_1 - ivE_2)\varphi(u, v)^t \, du \, dv$$

$$\varphi(u, v) = \varepsilon[\exp(2iuN^{-1/2}\cos 2\pi\mathbf{x}_1 . 2\mathbf{h} + 2ivN^{-1/2}\cos 2\pi\mathbf{x}_1 . \mathbf{h})]. \quad (A.6)$$

Then we obtain

$$\begin{aligned}\varphi(u, v) &= [1-(u^2/N)+(u^4/4N^2)] \times [1-(v^2/N)+(v^4/4N^2)] \\ &+ [2iu(N-1)^{-1}N^{-1/2}](E_{\mathbf{h}}^2-1) - iuv^2N^{-3/2} \\ &- iu^3N^{-5/2}(E_{\mathbf{h}}^2-1) - 2iuv^2N^{-5/2}(E_{\mathbf{h}}^2-1) \\ &+ (i/2)u^3v^2N^{-5/2} + (i/6)uv^4N^{-5/2} \\ &- (i/3)u^3N^{-5/2}(E_{3\mathbf{h}}^2-1) \\ &- [u^2/N(N-1)](E_{2\mathbf{h}}^2-1) \\ &- [v^2/N(N-1)](E_{\mathbf{h}}^2-1) + O(N^{-3}).\end{aligned} \quad (A.7)$$

Hence

$$\begin{aligned}\exp[(N/2)\ln\varphi(u, v)] &= \exp[(-u^2/2)-(v^2/2)]\{1+iuN^{-1/2}(E_{\mathbf{h}}^2-1) \\ &- iuv^2N^{-1/2}/2 - (u^2/2N)[(E_{2\mathbf{h}}^2-1)+(E_{\mathbf{h}}^2-1)^2] \\ &- u^4/8N - v^4/8N - u^2v^4/8N \\ &- (v^2/N)(E_{\mathbf{h}}^2-1) + O(N^{-3/2})\}.\end{aligned} \quad (A.8)$$

Using the formulae

$$H_n(x)\exp(-\tfrac{1}{2}x^2) = (2\pi)^{-1/2}\int_{-\infty}^{+\infty}(iu)^n \exp(-\tfrac{1}{2}u^2 - iux)\, du$$

$$(2\pi)^{-1/2}\int_{-\infty}^{+\infty}\exp(-x^2/2)H_n(x)H_m(x)\, dx = \delta_{nm}n!, \quad (A.9)$$

one gets

$$\begin{aligned}P(E_1, E_2) &\propto \exp(-\tfrac{1}{2}E_1^2 - \tfrac{1}{2}E_2^2) \\ &\times \{1+E_1(E_{\mathbf{h}}^2-1)N^{-1/2}+\tfrac{1}{2}E_1(E_2^2-1)N^{-1/2} \\ &+ [(E_1^2-1)/2N](E_{2\mathbf{h}}^2-1+(E_{\mathbf{h}}^2-1)^2) \\ &- H_4(E_1)/8N - H_4(E_2)/8N \\ &- (E_1^2-1)(E_2^2-1)/8N \\ &- (E_2^2-1)(E_{\mathbf{h}}^2-1)/N + O(N^{-3/2})\}.\end{aligned} \quad (A.10)$$

So we get

$$\begin{aligned}\varepsilon(\hat{E}_{2\mathbf{h}}|\hat{E}_{\mathbf{h}}) &= N^{-1/2}[(E_{\mathbf{h}}^2-1)+\tfrac{1}{2}(\hat{E}_{\mathbf{h}}^2-1)] \\ &\times \{1-[H_4(\hat{E}_{\mathbf{h}})/8N] \\ &- (\hat{E}_{\mathbf{h}}^2-1)(E_{\mathbf{h}}^2-1)/N\}^{-1} \\ &+ O(N^{-3/2}).\end{aligned} \quad (A.11)$$

So

$$\varepsilon(\hat{E}_{2\mathbf{h}}||\hat{E}_{\mathbf{h}}| = |E_{\mathbf{h}}|) \simeq \tfrac{3}{2}N^{-1/2}(E_{\mathbf{h}}^2-1) \quad (A.12)$$

and

$$\begin{aligned}\sigma^2(\hat{E}_{2\mathbf{h}}||\hat{E}_{\mathbf{h}}| = |E_{\mathbf{h}}|) &\simeq \sigma_u^2(\hat{E}_{2\mathbf{h}}||\hat{E}_{\mathbf{h}}| = |E_{\mathbf{h}}|) \\ &+ N^{-1}[E_{2\mathbf{h}}^2-1-(E_{\mathbf{h}}^2-1)^2] \\ &+ O(N^{-3/2}),\end{aligned} \quad (A.13)$$

where, up to the order N^{-1},

$$\sigma_u^2(\hat{E}_{2\mathbf{h}}||\hat{E}_{\mathbf{h}}| = |E_{\mathbf{h}}|) \simeq 1-(1/4N)(E_{\mathbf{h}}^2-1) - (1/4N)(E_{\mathbf{h}}^2-1)^2. \quad (A.14)$$

3. Derivation of (19) and (20)

Direct calculation gives

$$\begin{aligned}\langle(E_{\mathbf{k}}^2-1)(E_{2\mathbf{k}}^2-1)(E_{\mathbf{h}+\mathbf{k}}^2-1)\rangle_{\mathbf{k}} &\simeq N^{-1}[2N^{-1/2}E_{2\mathbf{h}}+4N^{-1}(E_{\mathbf{h}}^2-1) \\ &- 9N^{-3/2}E_{2\mathbf{h}}] \\ \langle(E_{\mathbf{k}}^2-1)^2(E_{2\mathbf{k}}^2-1)\rangle_{\mathbf{k}} &\simeq N^{-1}(6-5N^{-1}).\end{aligned} \quad (A.15)$$

Using the two formulae given by $(A.15)$, one derives immediately (19).

References

BARRA, J. R. (1971). *Notions Fondamentales de Statistique Matématique*, p. 244. Paris: Dunod.

BOURBAKI, N. (1961). *Eléments de Mathématique, Fonctions d'une Variable Réelle*, ch. 5. Paris: Hermann.

BROSIUS, J. (1979). Doctoraatsthesis K.U. Leuven, Belgium.

BUERGER, M. J. (1951). *Acta Cryst.* **4**, 531-544.

COCHRAN, W. (1954). *Acta Cryst.* **7**, 581-583.

GIACOVAZZO, C. (1976). *Acta Cryst.* **A32**, 967-976.

HAUPTMAN, H. (1964). *Acta Cryst.* **17**, 1421-1433.

HAUPTMAN, H. (1976). *Acta Cryst.* **A32**, 934-940.

HAUPTMAN, H. & KARLE, J. (1958). *Acta Cryst.* **11**, 149-157.

HEINERMAN, J. J. L., KRABBENDAM, H. & KROON, J. (1975). *Acta Cryst.* **A31**, 731-737.

KLUG, A. (1958). *Acta Cryst.* **11**, 515-543.